Factor Analysis Report for "Sleep Health and Lifestyle" Dataset

S/18/827 - J Kumudumali Sandaleka

April 7, 2024

1   Introduction

Factor analysis (FA) is a fundamental statistical technique used for modelling observed variables and their covariance structure in terms of a smaller number of underlying unobservable (latent) "factors". These factors aim to explain the underlying structure or patterns within a dataset by identifying common variance among observed variables.

In this study, the "Sleep Health and Lifestyle" dataset is used to conduct factor analysis. This dataset comprises various variables related to sleeping and lifestyle. By applying factor analysis, my objective is to reduce the dimensionality of the dataset while retaining meaningful information. This process facilitates the interpretation and the understanding of the underlying factors influencing sleep and lifestyle.

2   Methodology

The methodology section encompasses rigorous data preprocessing, exploratory factor analysis (EFA), and confirmatory factor analysis (CFA), ensuring a comprehensive understanding of latent factors and their relationships within the "Sleep Health and Lifestyle" dataset.

2.1   Dataset Description

The "Sleep Health and Lifestyle" dataset consists of 400 row and 13 columns, covering a range of variables including,

1. Person ID: An identifier for each individual.

2. Gender: The gender of the person (Male/Female).

3. Age: The age of the person in years.

4. Occupation: The occupation or profession of the person.

5. Sleep Duration (hours): The number of hours the person sleeps per day.

6. Quality of Sleep (scale: 1-10): A subjective rating of the quality of sleep.

7. Physical Activity Level (minutes/day): The number of minutes the person engages in physical activity daily.

8. Stress Level (scale: 1-10): A subjective rating of the stress level experienced by the person, ranging from 1 to 10.

9. BMI Category: The BMI category of the person (e.g., Underweight, Normal, Overweight).

10. Blood Pressure (systolic/diastolic): The blood pressure measurement of the person.

11. Heart Rate (bpm): The resting heart rate of the person in beats per minute.

12. Daily Steps: The number of steps the person takes per day.

13. Sleep Disorder: The presence or absence of a sleep disorder in the person (None, Insomnia, Sleep Apnea).

2.2 Factor Analysis

2.2.1 Data Preprocessing

Before conducting factor analysis, the dataset was thoroughly preprocessed to ensure its integrity and suitability. This required addressing duplicate entries and missing values in addition to standardizing the data to reduce inconsistencies caused by disparate units of measurement.In this stage I observed that there are no outliers in this dataset.

2.2.2 Exploratory Factor Analysis (EFA)

The preliminary stage of the analysis focused on determining the optimal number of factors to extract from the dataset. Various techniques were employed for this purpose, including Jolliffe's Method and the scree plot, which aimed to identify factors explaining 80% of the variance. EFA began with an initial factor analysis using Principal Component Analysis (PCA) or Maximum Likelihood Estimation (MLE) to establish loadings for each selected factor. To enhance interpretability, rotation techniques were applied.Finally, factor loading, commonalities, and uniqueness were interpreted for more understanding.

2.2.3 Confirmatory Factor Analysis (CFA)

Following the exploratory phase, CFA was used to determine how well the hypothesized model fit the observed data. This involved defining the correlations between the factors and the observed variables and evaluating the fit of the model using a variety of fit indices, such as the Tucker-Lewis Index (TLI), Comparative Fit Index (CFI), Root Mean Square Error of Approximation (RMSEA), and chi-square test. These indices validate the underlying factor structure by showing how well the suggested model fits the observed data.

In conclusion, the methodology included comprehensive data preprocessing, suitability assessment, exploratory factor analysis for factor extraction, and confirmatory factor analysis to validate the proposed model. These steps contribute to a thorough understanding of the dataset's latent factors and their relationships with observed variables.

3   Results and Discussion

In this section, we present the results of our analysis, beginning with an exploration of the basic assumptions underlying the factor analysis. We conducted Exploratory Factor Analysis (EFA) to identify underlying latent factors within the dataset, employing varimax rotation to enhance interpretability. The results of EFA provided insights into the structure of the data and potential underlying constructs. Following EFA, Confirmatory Factor Analysis (CFA) was performed to validate the factor structure identified in EFA.

3.1   Checking the basic assumptions

3.1.1   Sampling adequacy

The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy is a better measure of factor-ability. The minimum acceptable value is 0.50.

```
Kaiser-Meyer-Olkin factor adequacy
Call: KMO(r = standardized_data)
Overall MSA =  0.76
MSA for each item =
age           sleep_duration    quality_of_sleep    stress_level    bmi_category
0.77              0.85                0.71            0.73            0.83
systolic_bp       diastolic_bp      heart_rate          sleep_disorder
0.71              0.72                0.81            0.79
```

Initially, the KMO analysis for the "Sleep Health and Lifestyle" dataset yielded an overall Measure of Sampling Adequacy (MSA) of 0.59, which suggests marginal adequacy. After removing the variables with low MSA values, a second KMO analysis was conducted, resulting in a substantially improved overall MSA of 0.76. This indicates a higher level of sampling adequacy, suggesting that the remaining variables are more suitable for factor analysis.

3.1.2   Significance of the multiple correlations

This evaluates whether or not the variables intercorrelate at all, by evaluating the observed correlation matrix against an "identity matrix".

```
R was not square, finding R from data
$chisq              $p.value          $df
[1] 3709.504            0               36
```

The p-value of 0 indicates that the observed correlation matrix significantly differs from an identity matrix, as it is less than the conventional significance level of 0.05. Therefore, we reject the null hypothesis, suggesting that there are correlations among the variables in the dataset.

3.2 Explanatory Factor Analysis (EFA)

The Explanatory Factor Analysis (EFA) section delves into determining the appropriate number of factors to extract and estimating factor model parameters, revealing insights into underlying constructs within the dataset.

3.2.1 Determining the Number of Factors to Extract

After checking basic assumptions, we need to clarify that how many components or factors that we are going to extract. For this purposes, we can use some methods or criterion:

1. **Eigenvalues / Variances**

Factor	Eigen value	Proportion due to factor	Cumulative Proportion (%)
1	4.0555	0.4506	45.06
2	3.0713	0.3413	79.19
3	0.6944	0.0772	86.91
4	0.4476	0.0497	91.88
5	0.2858	0.0318	95.06
6	0.2507	0.0279	97.85
7	0.1176	0.0131	99.16
8	0.0511	0.0057	99.73
9	0.0206	0.0029	100.02

Based on the Jolliffe's method, it appears that the first two factors contribute significantly to the variance, with eigenvalues above 0.7. First three factors explains approximately 86.91% of the total variance, suggesting that they capture a substantial portion of the variability in the dataset.

2. **Scree Plot**
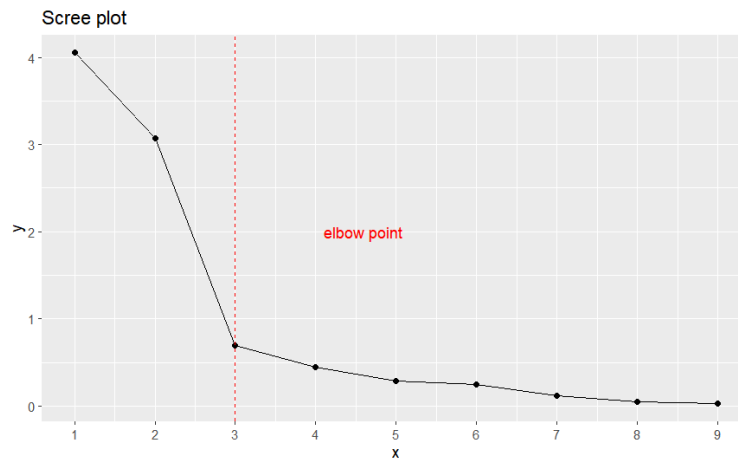
The scree plot shows the eigenvalues plotted against the factor number. The first three factors lie on the elbow.

Considering the findings of the scree plot and eigen values - variance we are positive that extracting three factors for our factor analysis of the "Sleep Health and Lifestyle" dataset is appropriate and meaningful.

3.2.2 Estimate the factor model parameters

In exploratory factor analysis (EFA), begin by conducting an initial factor analysis to determine the loadings for each selected factor. Two commonly used methods are Principal Component Analysis (PCA) and Maximum Likelihood Estimation (MLE):

	PCA Method			MLE Method		
Variable	**PC1**	**PC2**	**PC3**	**ML1**	**ML2**	**ML3**
Age	0.5	0.65	0.17	0.65	0.51	0.17
Sleep Duration	0.83	-0.03	-0.29	-0.03	0.85	-0.28
Quality of Sleep	0.97	0.03	-0.25	0.04	0.97	-0.25
Stress Level	-0.94	0.01	-0.01	0	-0.93	0.01
BMI Category	-0.21	0.66	0.45	0.62	-0.19	0.5
Systolic BP	-0.11	0.98	0.15	0.97	-0.11	0.19
Diastolic BP	-0.1	0.95	0.15	0.95	-0.1	0.19
Heart Rate	-0.69	0.23	0.01	0.22	-0.69	0.02
Sleep Disorder	-0.19	0.41	0.74	0.38	-0.19	0.77

In this analysis, both PCA and ML methods revealed that all three factors extracted from the analysis are statistically significant. The three factors together explain around 80% of the total variance in the data and factor loadings are also closer to -1 or +1, indicating stronger relationships between variables and factors in both methods. In terms of correlation patterns, PCA identifies factor 3 solely correlated with sleeping disorder, whereas ML shows factor 3 correlated with both sleeping disorder and BMI category.

Consequently, the Maximum Likelihood method is chosen for further analysis due to its more reliable factor loadings and better representation of the relationships between variables and factors.

3.2.3 Factor Rotations

Factor extraction aims to identify underlying constructs in the data, but these factors can be challenging to interpret directly. To address this issue, rotation techniques are used.

	Unrotated Factor Loadings			Rotated Factor Loadings		
Variable	**F1**	**F2**	**F3**	**F1**	**F2**	**F3**
Age	0.18	**0.8**	0.18	**0.65**	**0.51**	0.17
Sleep Duration	**-0.65**	**0.6**	-0.05	-0.03	**0.85**	-0.28
Quality of Sleep	**-0.67**	**0.74**	0	0.04	**0.97**	-0.25
Stress Level	**0.6**	**-0.67**	-0.22	0	**-0.93**	0.01
BMI Category	**0.71**	0.26	0.33	**0.62**	-0.19	**0.5**
Systolic BP	**0.81**	**0.58**	-0.01	**0.97**	-0.11	0.19
Diastolic BP	**0.79**	**0.57**	-0.01	**0.95**	-0.1	0.19
Heart Rate	**0.6**	-0.35	-0.19	0.22	**-0.69**	0.02
Sleep Disorder	**0.61**	0.08	**0.62**	0.38	-0.19	**0.77**

According to the results, the rotated factor loadings demonstrate notable correlations closer to 1 and -1, indicating stronger and more meaningful relationships between variables and factors. Given these findings, it is sound to continue with factors rotated using varimax rotation.

3.2.4 Varimax - MLE Factor Model

Variable	Estimated Factor Loadings			Communalities	Specific Variances
	F1	**F2**	**F3**	h_i^2	$\psi = 1 - h^2$
Age	**0.65**	**0.51**	0.17	0.7084155	0.291584545
Sleep Duration	-0.03	**0.85**	-0.28	0.7917550	0.208245049
Quality of Sleep	0.04	**0.97**	-0.25	0.9950148	0.004985194
Stress Level	0.00	**-0.93**	0.01	0.8636188	0.136381241
BMI Category	0.62	-0.19	**0.50**	0.6716989	0.328301116
Systolic BP	**0.97**	-0.11	0.19	0.9950032	0.004996797
Diastolic BP	**0.95**	-0.10	0.19	0.9512039	0.048796081
Heart Rate	0.22	**-0.69**	0.02	0.5186859 *	0.481314104
Sleep Disorder	0.38	-0.19	**0.77**	0.7663940	0.233606008
Eigen Values	4.055	3.071	0.694		

Factor Analysis
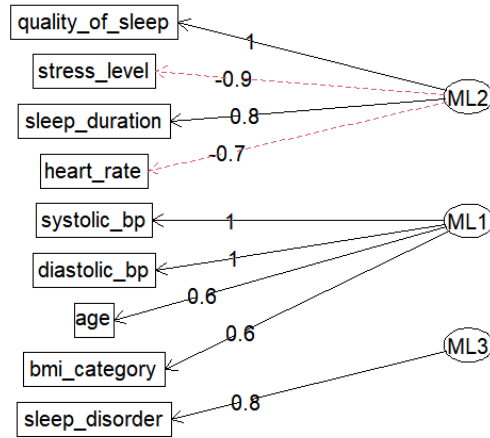
Based on the factor loadings provided in the table, the following improvements and factor names could be suggested:

1. **Factor Loadings**

   - Factor 1 shows strong positive correlations with Systolic Blood Pressure (0.97) and Diastolic Blood Pressure (0.95), indicating that it represents aspects related to blood pressure. Moreover, a moderate positive correlation with Age (0.65) suggests that older individuals tend to have higher blood pressure.

   - Factor 2 exhibits a strong positive correlation with Quality of Sleep (0.97) and a strong negative correlation with Stress Level (-0.93), suggesting that higher levels of stress are associated with lower quality of sleep. Additionally, positive correlations with Sleep Duration (0.85) and Age (0.51) imply that older individuals with longer sleep durations experience higher quality of sleep. The negative correlation with Heart Rate (-0.69) may indicate that lower heart rates are associated with better sleep quality.

   - Factor 3 is positively correlated with BMI category (0.72), indicating that it represents aspects related to body composition. Conversely, it shows a negative correlation with Sleep Disorder (-0.47), suggesting that lower scores on this factor are associated with fewer sleep disorders.

2. **Communalities**

Another assessment of how well this model performs can be obtained from the communalities. The model performs particularly well in explaining the variation in Quality of Sleep and Systolic Blood Pressure, while also showing relatively strong performance for variables such as Diastolic Blood Pressure, Stress Level, Sleep Duration, Age, Sleep Disorder, and BMI Category. However, the model performs less effectively for Heart Rate, explaining half of the variation.

3. **Total communality value**

The sum of all communality values is the total communality value = 7.261763.

4. **The proportion of total variation explained by the three factors**

$$\frac{7.261763}{9} = 0.8068625$$

Therefore, 80.69% proportion of the total variation explained by the three factors.

5. **Uniqueness**

The uniqueness of a variable represents the variance that is unique to that variable and not shared with other variables. According to this analysis, the uniqueness of Heart Rate is 0.48, it means that 48% of the variance in Heart Rate is not shared with other variables in the factor model. Conversely, Quality of Sleep and Systolic BP indicate that only 0.4% of their variance is not accounted for by other variables. It is important to note that, variables with lower uniqueness values are more important in explaining the common variance captured by the factors.

3.3 Confirmatory Factor Analysis (CFA)

CFA is used to assess the overall measurement of a concept when there are multiple items available to measure it.

```
Test statistic                                    314.838
Degrees of freedom                                     23
P-value (Chi-square)                               0.000


User Model versus Baseline Model:

Comparative Fit Index (CFI)                        0.922
Tucker-Lewis Index (TLI)                           0.877


Root Mean Square Error of Approximation:

RMSEA                                              0.184
90 Percent confidence interval - lower             0.166
90 Percent confidence interval - upper             0.203
P-value H_0: RMSEA <= 0.050                        0.000
P-value H_0: RMSEA >= 0.080                         1.000


Standardized Root Mean Square Residual:

SRMR                                               0.071


R^2 Values
age           quality_of_sleep    stress_level  sleep_duration  heart_rate
0.675                       NA           0.807           0.779       0.435
systolic_bp        diastolic_bp    bmi_category  sleep_disorder
0.992                    0.954           0.796           0.526
```

3.3.1 Overall Model Level Fit

Multiple methods of estimation have been developed for SEM models generally and CFA models specifically.

1. **Model Test:**

   The chi-square test assesses the goodness of fit of the model. A lower p-value (close to 0) suggests that the model significantly deviates from the expected values, indicating a potential lack of fit. In this case, with a p-value of 0.000, the model significantly deviates from the expected values.

2. **Root Mean Square Error of Approximation (RMSEA):**

   The root mean square error of approximation is a parsimonyadjusted fit index, meaning that it favors simplicity in models. Lower RMSEA values indicate better fit, with values below 0.05 typically considered good. The provided RMSEA of 0.184 suggests a moderate fit.

3. **Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI):**

These indices assess the model fit by comparing it to a baseline model. Values closer to 1 indicate better fit. Both CFI and TLI values are closer to 1, suggesting a reasonable fit of the model.

4. **Standardized Root Mean Square Residual (SRMR):**

SRMR measures s the square root of the standardized difference between the sample covariances and the covariances predicted by the model. Lower values indicate better fit. The provided SRMR of 0.071 suggests a reasonably good fit.

Overall, the model appears to have a reasonable fit to the data, although it may not be perfect. Further adjustments or evaluations might be necessary depending on the specific goals of the analysis.

3.3.2 Equation Level Fit

The equation-level fit of the model was assessed using R^2 values for each observed variable. The model demonstrated strong predictive power for variables such as stress level, sleep duration, systolic and diastolic blood pressure, and BMI category, with R^2 values ranging from approximately 79.6% to nearly 99.2%. Variables like age and sleep disorder also exhibited moderate to high levels of explained variance (approximately 67.5% and 52.6% respectively). Notably, heart rate showed comparatively lower explained variance at around 43.5%. These results suggest that the model provides a generally good fit for predicting the observed variables.

3.3.3 Parameter Level Fit

Factor loadings indicate the strength and direction of the relationship between latent factors and observed variables.

	λ			θ
	Factor 1	**Factor 2**	**Factor 3**	**Error**
Age	0.555	0.676	0.000	0.325
Sleep Duration	1.000	0.000	0.000	-0.001
Quality of Sleep	-0.898	0.000	0.000	0.193
Stress Level	0.883	0.000	0.000	0.221
BMI Category	-0.659	0.000	0.000	0.565
Systolic BP	0.000	0.996	0.000	0.008
Diastolic BP	0.000	0.977	0.000	0.046
Heart Rate	0.000	0.000	0.892	0.204
Sleep Disorder	0.000	0.000	0.725	0.474

In Factor 1, variables like age, sleep duration, and systolic blood pressure have relatively high loadings, suggesting a strong association with the latent factor represented by Factor 1. Conversely, variables like stress level and heart rate show negative loadings, indicating an inverse relationship with Factor 1.

Factor 2 primarily captures variables related to blood pressure (systolic and diastolic), with high loadings on these variables and relatively low loadings on others.

Factor 3 is predominantly associated with BMI category and sleep disorder, with high loadings on these variables and negligible loadings on others.

```
$psi
factr1 factr2 factr3
factor1  1.000
factor2 -0.120  1.000
factor3 -0.361  0.795  1.000
```

Theta and Psi matrices represent the relationships between factors and error terms. These values indicate the strength of association and covariance between latent factors and any unexplained variance in the observed variables.

Overall, the factor loadings suggest that the model adequately captures the relationships between latent factors and observed variables, providing insights into the underlying structure of the data and informing potential model refinement. Non-significant loadings, such as those for certain variables on Factor 3, may indicate areas for model improvement, possibly through the inclusion of additional variables or modification of the structural model.

4 Conclusion

In conclusion, the Explanatory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) conducted on the "Sleep Health and Lifestyle" dataset have revealed three distinct factors that explains various facts of sleep health and overall well-being. Factor 1, identified as the Cardiovascular Health Factor, contains variables related to blood pressure and age, highlighting the critical role of cardiovascular health in sleep quality. Factor 2, named the Sleep Wellness Factor, includes variables associated with sleep duration, quality, and stress levels, highlighting relationship between sleep patterns and psychological well-being. Finally, Factor 3, the Body-Mind Balance Factor, reflects variables related to body composition and sleep disorders, emphasizing the inter connection of physical and mental health with sleep quality.

These factors offer valuable insights for healthcare professionals, researchers, and policymakers looking to create specific measures to improve sleep quality and promote overall well-being. Understanding the different dimensions of sleep health and lifestyle factors allows for the development of customized solutions to address the specific needs and challenges that individuals in various populations face. Furthermore, the discovery of these factors emphasizes the importance of taking a variety of perspectives when it comes to sleep health management.

5 References

1. CitedDemir, Ergul. "RPubs - Intro to EFA and CFA in R." Rpubs.com, 2022, `rpubs.com/isbell_daniel/efa_cfa_intro`. Accessed 7 Apr. 2024.

2. Example of Factor Analysis Method Section Reporting. Huang, Gaoping. "Exploratory Factor Analysis – Notes and R Code · Gaoping Huang's Blog." Gaopinghuang0.Github.io, 9 Feb. 2018, `gaopinghuang0.github.io/2018/02/09/exploratory-factor-analysis-notes-and-R-code`. Accessed 23 Nov. 2022.

3. Isbell, Dan. "RPubs - MyClassNotes_EFA_CFA." Rpubs.com, 17 Apr. 2021, `rpubs.com/Symrna/EFA_CFA`.

4. Murphy, Phil. "RPubs - Exploratory Factor Analysis in R." Rpubs.com, 21 Apr. 2021, `rpubs.com/pjmurphy/758265`.

5. THARMALINGAM, LAKSIKA. "Sleep Health and Lifestyle Dataset." `www.kaggle.com`, Oct. 2023, `https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset`.

6 Appendices

```
1   ---
2   title: "ST 405 Assignment 1"
3   author: "S18827"
4   date: "2024-04-06"
5   output: pdf_document
6   ---
7   ```{r}
8   # Load necessary libraries
9   library(psych)  # for EFA
10  library(lavaan) # for CFA
11  library(tidyverse)
12  library(corrplot)
13  library(janitor)
14  library(semPlot)
15  ```
16
17  ###Pre - processing data
18
19  ```{r}
20
21  life_exp_data <- read_csv("Sleep_health_and_lifestyle_dataset2.csv",
22                            col_types = 'cfifdiiifciif')
23
24
25  life_exp_data <- distinct(life_exp_data)
26
27  # Define factor column names
28  factor_column_names <- c("Gender", "Sleep Disorder")
29
30  # Check factor levels for multiple columns
31  lapply(life_exp_data[, c("Gender","BMI Category", "Sleep Disorder")], levels)
32
33  |
34
35  # Convert factor columns to numeric
36  life_exp_data[, factor_column_names] <- lapply(life_exp_data[, factor_column_names], as.numeric)
37
```

```r
38    # Separate Blood Pressure into Systolic and Diastolic columns
39    life_exp_data <- life_exp_data %>%
40      separate(col = "Blood Pressure", into = c("Systolic_bp", "Diastolic_bp"), sep = "/") %>%
41      mutate(across(c(Systolic_bp, Diastolic_bp), as.integer)) %>%
42      mutate(life_exp_data,
43            `BMI Category` = recode(`BMI Category`,"Under Weight" = 1,
44                                              "Normal" = 2,
45                                              "Overweight" = 3,
46                                              "Obese" = 4))
47
48    life_exp_data <- clean_names(dat = life_exp_data, case = "snake")
49
50    # Exclude non-numerical variables and person_id
51    boxplot(life_exp_data[, !(names(life_exp_data) %in% c("blood_pressure", "occupation", "person_id"))])
52
53
54    # Check the structure of the dataset
55    str(life_exp_data)
56
57
58    ```
59
60    ```{r}
61    # Exclude non-numeric columns
62    numeric_data <- life_exp_data[, sapply(life_exp_data, is.numeric)]
63
64    standardized_data <- scale(numeric_data)
65    # Compute KMO
66    kmo_result <- KMO(standardized_data)
67    print(kmo_result)
68    ```
69
70    ```{r}
71    standardized_data <- standardized_data[, KMO(standardized_data)$MSAi > 0.5] # Get rid of all
        variables with MSA < 0.50
72    kmo_result <- KMO(standardized_data)
73    print(kmo_result)
74    ```
```

```r
76    ```{r}
77    cortest.bartlett(standardized_data)
78    ```
79
80    ```{r}
81
82    # Compute the correlation matrix
83    correlation_matrix <- cor(standardized_data)
84
85    # Create a correlation plot
86    corrplot(correlation_matrix, method = "circle", type = "upper", tl.col = "black")
87
88    ```
89
90    ```{r}
91    cov_mat <- cov(standardized_data)
92    round(cov_mat, 2)
93    ```
94
95    ###Explonatory Factor Analysis
96
97    ```{r}
98    eigen_Val <- round(eigen(correlation_matrix)$values,4)
99    no_factors <- length(eigen_Val)
100   Proportion_due_to_factor <- round(eigen_Val/sum(eigen_Val),4)
101
102   data.frame(Factor = 1:no_factors,
103             Eigen_value = eigen_Val,
104             Proportion_due_to_factor,
105             Cummulative_Proportion = round(cumsum(Proportion_due_to_factor)*100, digits = 2))
106   ```
107
108   ```{r}
109   n <- dim(correlation_matrix)[1]
110   scree_tb <- tibble(x = 1:n,
111                      y = sort(eigen(correlation_matrix)$value, decreasing = TRUE))
112
113   scree_plot <- scree_tb %>%
```

```r
113   scree_plot <- scree_tb %>%
114     ggplot(aes(x, y)) +
115     geom_point() +
116     geom_line()  +
117     scale_x_continuous(breaks = 1:n) +
118     ggtitle("Scree plot")
119
120   scree_plot +
121     geom_vline(xintercept = 3, color = "red", linetype = "dashed") +
122     annotate("text", 4.1, 2, label = "elbow point", color = "red", hjust = 0)
123   ```
124
125
126   ```{r}
127   # Perform Exploratory Factor Analysis
128   fa_varimax_pca <- fa(standardized_data, nfactors = 3,
129                        rotate = "varimax",
130                        fm = "pa")
131   fa_varimax_pca
132
133   # Extract factor loadings
134   load <- fa_varimax_pca$loadings
135
136   fa.diagram(load, main = "Principal Component Method")
137   ```
138
139   ```{r}
140   # Perform Exploratory Factor Analysis
141   fa_variax_ml <- fa(standardized_data, nfactors = 3,
142                      rotate = "varimax",
143                      fm = "ml")
144   fa_variax_ml
145
146   # Extract factor loadings
147   load <- fa_variax_ml$loadings
148
149   fa.diagram(load, main = "Maximum Likelihood Estimation")
150   ```
```

```r
{r}
# Perform Exploratory Factor Analysis
fa_no_rotation_ml <- fa(standardized_data, nfactors = 3,
                        rotate = "none",
                        fm = "ml")
fa_no_rotation_ml

# Extract factor loadings
load <- fa_no_rotation_ml$loadings

fa.diagram(load)
```

```r
{r}
fa_variax_ml$e.values[1:3]
```

```r
{r}
sum(fa_variax_ml$communalities)
sum(fa_variax_ml$communalities)/9
```

###Confirmatory Factor Analysis

```r
{r}
#Model Identification

p <- ncol(standardized_data)
q <- 3 * p   # Assuming each observed variable loads on all 3 factors

# Check if q ≤ p(p + 1)/2 for identification
identification_status <- q <= p * (p + 1) / 2

# Print identification status
if (identification_status) {
  print("CFA model is identified.")
} else {
  print("CFA model is not identified. Adjust the model to meet identification criteria.")
```

```r
{r}
cfa_model <- '
factor1 =~ age + quality_of_sleep + stress_level + sleep_duration + heart_rate
factor2 =~  systolic_bp + diastolic_bp + age
factor3 =~ bmi_category + sleep_disorder'

Fit <- cfa(cfa_model, data = standardized_data)
summary(Fit)
```

```r
{r}
summary(Fit, fit.measures = T)
```

```r
{r}
summary(Fit, standardized = T)
```

```r
{r}
#standardize Loadings
inspect(Fit, what = "std")

inspect(Fit, what = "r2")
```

```r
{r}
# Generate the SEM plot
semPaths(Fit,
         layout = "tree",  # Use tree layout
         whatLabels = "est",  # Show parameter estimates
         rotation = 3
         )

semPaths(Fit, "std")
```